a. Using Python with BeautifulSoup, extract the data from the Web site and create a csv dataset containing that data.

- Please check out "TOP500ListJune2018.csv"

b. Clean & explore the dataset, producing summary statistics for Cores, RMax, RPeak, and Power.

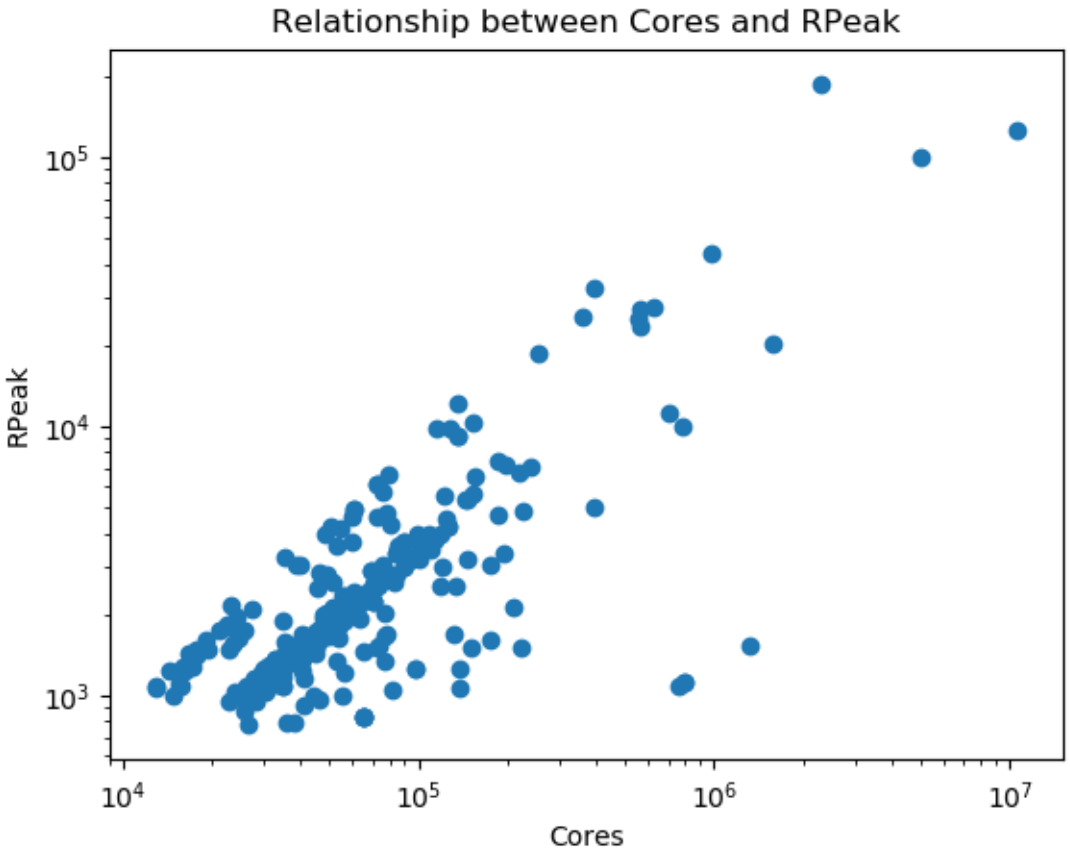
- After dropping rows containing missing values, there are 264 rows. The following summary statistics are based on these rows.

	Cores	RMax	RPeak	Power
count	2.640000e+02	264.000000	264.000000	264.000000
mean	1.711464e+05	3168.896591	4870.415152	1596.928030
std	7.480408e+05	10356.172212	15650.198717	2545.841624
min	1.288000e+04	715.600000	778.500000	47.000000
25%	3.656000e+04	852.775000	1342.375000	486.000000
50%	5.332600e+04	1074.150000	1822.550000	840.000000
75%	8.989200e+04	2335.875000	3207.900000	1543.250000
max	1.064960e+07	122300.000000	187659.300000	19431.000000

c. Display and explain the relationship between

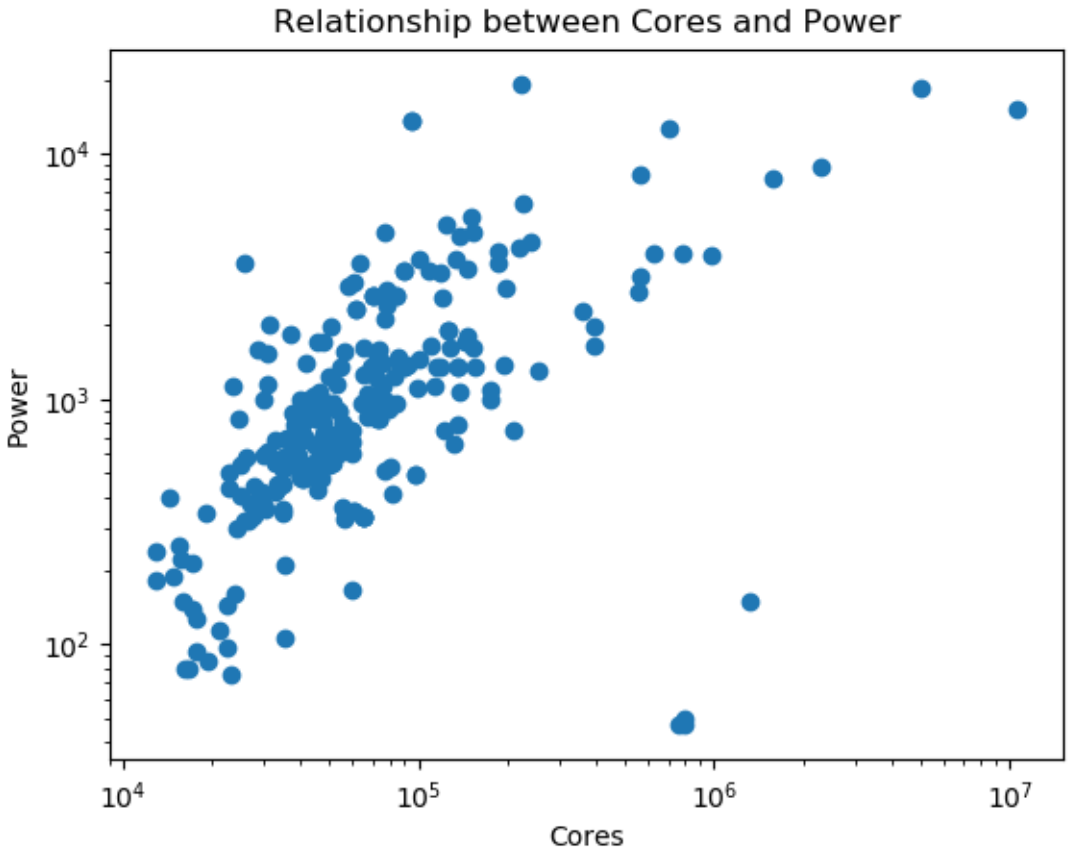
a. Cores and RPeak

- From the following figure, I can see Cores and RPeak have positive linear relationship on log-log axes.



b. Cores and Power

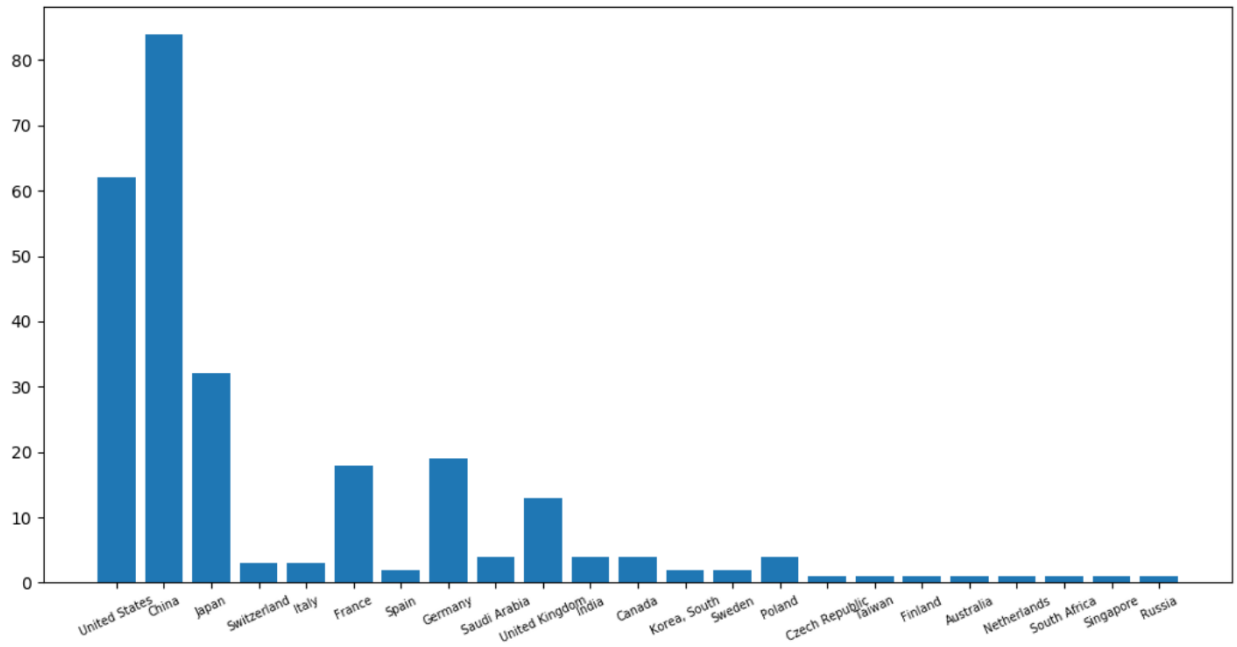
- From the following figure, I can see Cores and Power have positive linear relationship on log-log axes.



Display summary statistics and visualizations for the Country data.

- Following is the count of frequency for each country (Based on "TOP500ListJune2018.csv" and dropped rows with missing values in any columns)

Counter({'China': 84, 'United States': 62, 'Japan': 32, 'Germany': 19, 'France': 18, 'United Kingdom': 13, 'Saudi Arabia': 4, 'India': 4, 'Canada': 4, 'Poland': 4, 'Switzerland': 3, 'Italy': 3, 'Spain': 2, 'Korea, South': 2, 'Sweden': 2, 'Czech Republic': 1, 'Taiwan': 1, 'Finland': 1, 'Australia': 1, 'Netherlands': 1, 'South Africa': 1, 'Singapore': 1, 'Russia': 1})


80
70
60
50
40
30
20
10
0
United States
China
Japan
Switzerland
Italy
France
Spain
Germany
Saudi Arabia
United Kingdom
India
Canada
Korea, South
Sweden
Poland
Czech Republic
Taiwan
Finland
Australia
Netherlands
South Africa
Singapore
Russia